



Steekproefmethoden

Prof. dr. Herman Callaert

Er zijn zo goed als geen realistische studies waarbij je de volledige populatie kan opmeten. Daarom verzamel je informatie over een deel van de populatie, namelijk over de elementen in een steekproef. Er zijn heel veel manieren om een steekproef te trekken en er zijn ook heel wat verschillende manieren om eigenschappen van de getrokken elementen op te meten. In deze tekst maak je kennis met enkele basisbegrippen. Zij worden geïllustreerd met “echt gebeurde” verhalen.

De structuur van deze tekst is als volgt.

- In de eerste twee paragrafen vind je een korte motivatie waarom je met steekproeven werkt. Bij de basisterminologie krijgt het begrip “steekproeflijst” extra aandacht.

De volgende twee paragrafen beschrijven hoe je de data verzamelt. Dat gebeurt in twee stappen.

- In paragraaf 3 leer je dat er verschillende manieren zijn om aan de elementen van een steekproef te geraken. Hier hoort ook een eigen benaming bij.
- Nadat je de elementen van je steekproef verzameld hebt moet je bij die elementen de karakteristieken opmeten die nuttig zijn voor je onderzoek. Je kan hierbij twee basismethoden onderscheiden. Zij worden beschreven in paragraaf 4.

Bij het verzamelen van de steekprofelementen en bij het opmeten van karakteristieken kan er heel veel fout gaan. Je vindt hiervan voorbeelden in paragraaf 5 en 6.

- Paragraaf 5 beschrijft wat er zoal mis kan lopen bij het verzamelen van de steekprofelementen.
- In paragraaf 6 kan je lezen waar je allemaal moet op letten bij het opmeten van de gegevens.

Inhoudstafel

1.	Waarom met een steekproef werken?	3
1.1.	De volledige populatie testen is te duur en het duurt te lang.	3
1.2.	De volledige populatie is er nog niet.....	3
1.3.	Testen vernietigt de populatie.	3
2.	Terminologie	4
2.1.	De steekproeflijst.....	4
3.	Methoden om aan de elementen van een steekproef te geraken	5
3.1.	Niet-toevalsgestuurde methoden (selecte steekproeven)	5
3.1.1.	Steekproef op basis van vrijwillige reactie (<i>voluntary respons sampling</i>)	5
3.1.2.	Steekproef op basis van gemakkelijke bereikbaarheid (op goed geluk af; in het wilde weg) (<i>accessibility sampling, convenience sampling</i>)	5
3.1.3.	Steekproef op basis van quota (<i>quota sampling, judgment sampling</i>).....	6
3.2.	Toevalsgestuurde methoden (aselecte steekproeven).....	6
3.2.1.	EAS Enkelvoudige aselecte steekproef (<i>SRS simple random sampling</i>)	6
3.2.2.	Gestratificeerde steekproef (<i>Stratified random sampling</i>).....	7
3.2.3.	Geclusterde steekproef (<i>Cluster sampling</i>)	8
3.2.4.	Systematische steekproef (<i>Systematic sampling</i>).....	9
3.2.5.	Getrapte steekproef (<i>multi-stage sampling</i>)	10
4.	Methoden om eigenschappen van steekproefelementen op te meten	10
4.1.	De “rechtstreekse” observatie of meting.....	10
4.2.	De enquête (<i>survey</i>).....	11
5.	Vertekening door de manier waarop je aan de elementen van je steekproef geraakt (<i>sample selection bias</i>)	11
5.1.	Vertekening door de selectiemethode die niet “neutraal” is (niet gebaseerd op een vooraf gekend toevalsmechanisme)	11
5.1.1.	Steekproef op basis van vrijwillige reactie	11
5.1.2.	Steekproef op basis van bereikbaarheid	12
5.1.3.	Steekproef op basis van quota	12
5.2.	Vertekening door een onvolledige of een verkeerde steekproeflijst	13
5.2.1.	Ondervertegenwoordiging (<i>undercoverage</i>)	13
5.2.2.	Vertekening door omvang (<i>size bias</i>).....	13
5.3.	Vertekening door ontbrekende respons	14
5.3.1.	Ontbrekende informatie bij directe opmetingen (<i>missing data</i>).....	14
5.3.2.	Non-respons bij enquêtes (<i>non-response bias</i>)	15
6.	Vertekening door de manier waarop je informatie opmeet of inwint (<i>response bias</i>).....	15
6.1.	Vertekening door het vragenformulier van de enquête (<i>questionnaire bias</i>).....	15
6.1.1.	Vragen die onnauwkeurig of te ingewikkeld zijn (<i>wording of questions</i>)	15
6.1.2.	Vragen die gevoelig liggen (<i>sensitive questions</i>).....	16
6.2.	Vertekening door foutief antwoord of door slecht geijkt meetinstrument	16
6.2.1.	Systematische meetfouten (<i>measurement bias</i>)	16
6.2.2.	Verkeerd antwoord (<i>incorrect response bias</i>).....	16

1. Waarom met een steekproef werken?

Er zijn veel redenen waarom je met een steekproef werkt, en die redenen zijn dikwijls heel concrete dingen zoals tijd, geld, geschoolde medewerkers enz. Statistiek is inderdaad een wetenschap die heel veel met de realiteit te maken heeft.

1.1. *De volledige populatie testen is te duur en het duurt te lang.*

De “Volkstelling” (tot 1991) of de “Algemene socio-economische enquête” (2001) werd in België om de 10 jaar gehouden. Daarbij werd elk gezin bevestigd. Het opstellen van een efficiënte vragenlijst (met respect voor de privacy) en het afnemen van de enquête (drukken van de formulieren, verzenden, herinnering sturen, ...) is enorm complex en vraagt zeer veel tijd en mankracht. Het verwerken van de gegevens duurt nadien nog jaren. Het gebruik van dit logge instrument is in bijna alle gevallen veel te duur en veel te tijdrovend. Wat doe je bijvoorbeeld als je de jaarlijkse evolutie van het woon-werkverkeer nodig hebt om beslissingen over het openbaar vervoer snel en efficiënt bij te sturen?

De Verenigde Staten is een groot land met een complexe bevolkingsstructuur. Het is geweten dat er daar bij een volkstelling heel wat mensen “gemist” worden, terwijl er sommigen dubbel worden geteld. Universiteitsstudenten die ver van huis een studio kopen om daar enkele jaren te wonen worden op die plaats meegeteld. Maar als hun ouders die naam ook nog eens opschrijven als zij de samenstelling van hun gezin opgeven, wordt zo’n student dubbel geteld. Anderzijds worden specifieke groepen zoals daklozen, analfabeten, anderstalige inwijkelingen, ... (gedeeltelijk) gemist. In een bepaalde verkrotte en onveilige wijk van een grote stad waren er bij een volkstelling veel minder huizen opgetekend dan dat er voor elektriciteitstellers (per huis) betaald werd!

1.2. *De volledige populatie is er nog niet.*

De leef- en eetgewoonten van de groep vrouwen tussen 20 en 40 zijn vandaag in Vlaanderen niet drastisch verschillend van de manier waarop vrouwen tussen 20 en 40 vijf jaar geleden leefden. Als je mag onderstellen dat er ook de komende vijf jaar geen grote verschillen zullen optreden, dan is het zinvol om te kijken naar het geboortegewicht van baby’s die in Vlaanderen geboren worden in een tijdsspanne van 10 jaar (de 5 voorbije jaren en de 5 komende jaren). Je kan bijvoorbeeld iets willen weten over het percent baby’s met een te laag geboortegewicht. Maar de helft van die baby’s is er nog niet!

1.3. *Testen vernietigt de populatie.*

Bij het proeven van wijn schenkt men wat wijn in een glas, en dan proeft je een slokje (voor de echte wijnkenners gaat dat met een hele ceremonie gepaard). En dan zeg je: ja, die fles is lekker. Maar dat kan je natuurlijk langs geen kanten weten. Is de rest van de wijn in je glas even lekker? Zekerheid heb je enkel als je dat ook proeft. En is elk volgend glas uit die fles even lekker? Dat weet je enkel als je elk volgend glas ook proeft. En nadat je de volledige populatie hebt getest is de fles leeg.

Ter info. Als je een eerste (ietwat oppervlakkige) definitie wil hanteren dan zou je kunnen zeggen dat wijn proeven “exploratieve statistiek” is (de slokjes wijn die jij hebt geproefd zijn goed bevonden). Als je op basis daarvan zegt dat de hele fles lekker is, dan doe je aan “verklarende statistiek”.

2. Terminologie

populatie (*population*)

de totale verzameling elementen waarover je relevante informatie wil bekomen

steekproeflijst of steekproefkader (*frame*)

de volledige lijst van elementen waarover je beschikt en waaruit je een steekproef trekt.

steekproef (*sample*)

de verzameling elementen die je uit de steekproeflijst hebt gehaald

2.1. De steekproeflijst

“tussen populatie en steekproef staan lijsten in de weg en praktische bezwaren”

Als je aan genoeg voorbeelden uit de realiteit denkt dan ontdek je dat het trekken van een steekproef niet eenvoudig is. In theorie zou dat geen probleem mogen zijn. Je hoeft immers alleen maar alle elementen uit je populatie een nummer te geven. En dan trek je uit die lijst volgens een of andere steekproefmethode.

Als je een steekproef uit alle leerlingen van je school wil trekken dan kan je ze inderdaad allemaal een volgnummer geven. Maar hoe maak je een lijst van alle internetgebruikers of van alle beukenbladeren in het Zoniënwoud?

Een steekproef kan je alleen maar trekken uit een lijst die je kent. Je moet dus proberen om die lijst zo te maken dat hij zo goed mogelijk bij de populatie aansluit. Je statistische uitspraken zijn immers veralgemeningen van je steekproef naar je steekproeflijst. En als die lijst op belangrijke punten verschilt van de te onderzoeken populatie, dan heb je een probleem. Met statistiek kan je het probleem van een slechte steekproeflijst achteraf niet opvangen.

In het ideale geval valt een steekproeflijst samen met de bestudeerde populatie, maar in de praktijk kan hij te veel elementen, te weinig elementen, of beide bevatten. Bij verkiezingen in België worden kiezerslijsten opgemaakt van iedereen die verplicht is om te gaan stemmen. Mensen die recent van de ene gemeente naar de andere zijn verhuisd komen soms door een administratieve vergissing niet op zo’n kiezerslijst voor. De mensen die in de periode tussen het opmaken van de lijst en de dag van de verkiezingen overleden zijn staan daar dan wel nog op.

3. Methoden om aan de elementen van een steekproef te geraken

Slechte methoden om een steekproef te trekken leveren onbetrouwbare gegevens waarmee je geen statistisch verantwoorde uitspraken kan doen. Als je rapporten leest die gebaseerd zijn op gebrekkige steekproefmethoden, wees dan ten zeerste op je hoede voor de besluiten die daar getrokken worden.

3.1. Niet-toevalsgestuurde methoden (selecte steekproeven)

3.1.1. Steekproef op basis van vrijwillige reactie (*voluntary respons sampling*)

Bij een steekproef op basis van vrijwillige reactie wordt een grote groep mensen uitgenodigd om te antwoorden. Al wie antwoordt wordt meegeteld in de steekproef.

Een columnist van een grote krant vraagt zich af of ouders met kinderen die nog zouden willen als zij de kans kregen om het over te doen. Zij roept in haar krant ouders op om hun antwoord naar de redactie te sturen.

Op de radio loopt een live programma over een controversieel onderwerp en de luisteraars worden opgeroepen om te bellen en hun mening te geven.

Een operator wil weten of mensen bereid zijn extra te betalen voor digitale televisie. Daarom organiseert hij op zijn website een zogenaamde “internetpoll”, waar je je mening kan aanklikken en dan stemmen.

Een steekproef op basis van vrijwillige reactie is geen goede methode voor statistisch onderzoek. Je hebt geen enkel zicht op de manier waarop de steekproef tot jou komt en dikwijls kan je zelfs geen nauwkeurige beschrijving geven van de populatie waaruit deze steekproef komt. Er treden allerlei soorten vertekening op. Mensen die kwaad zijn reageren sneller dan mensen die tevreden zijn. Mensen met een uitgesproken mening zullen sneller een radioprogramma opbellen dan anderen. Op een internetenquête kan eenzelfde persoon 10 keer antwoorden, als hij tenminste internetverbinding heeft en de juiste website kan vinden.

3.1.2. Steekproef op basis van gemakkelijke bereikbaarheid (op goed geluk af; in het wilde weg) (*accessibility sampling, convenience sampling*)

Bij een steekproef op basis van bereikbaarheid neem je die mensen (of elementen) waar je het gemakkelijkste bij kan komen, of die je zonder veel extra inspanning toevallig tegenkomt. Je wil weten of de leerlingen van je school vinden dat roken ongezond is en daarom vraag je dat gewoon aan je vrienden in je klas.

Om te weten of de mensen tevreden zijn over de kwaliteit van de aangeboden producten stelt een verantwoordelijke van de supermarkt op zaterdagvoormiddag die vraag aan de eerste 250 klanten die komen winkelen.

Een steekproef op basis van bereikbaarheid is geen goede methode omdat je helemaal niet weet of de elementen in je steekproef een goed beeld geven van de populatie. Je kan er bijna zeker van zijn dat dit niet zo is. Je vrienden zijn misschien juist je vrienden omdat zij denken zoals jij denkt. Als je alleen maar hun mening vraagt dan geeft dat geen goed beeld van de hele school. En als de baas van die supermarkt wil weten wat de mensen over zijn producten denken, dan volstaat het niet om alleen maar de mening te horen van wie op zaterdagochtend

vroeg komt winkelen. Bovendien neemt hij een steekproef uit zijn eigen klanten en hoort hij de mening niet van mensen die uit ontevredenheid al maanden niet meer naar zijn supermarkt komen.

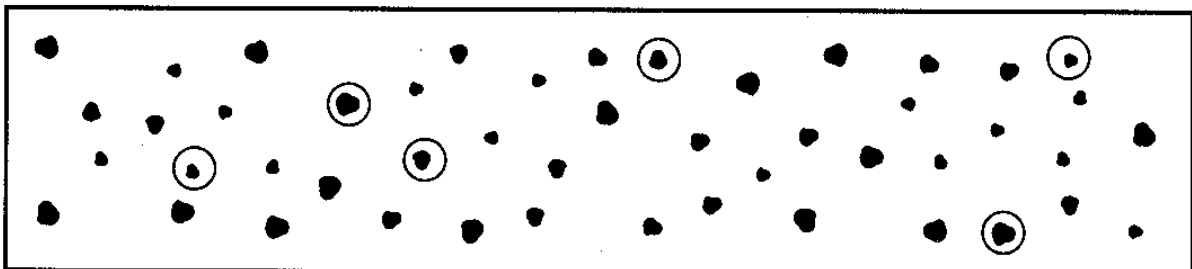
3.1.3. Steekproef op basis van quota (*quota sampling, judgment sampling*)

Om een steekproef te bekomen die een goede weergave is van de populatie kan je je laten leiden door het oordeel van een expert die zeer veel over de populatie weet. Die expert zal dan proberen om de populatie “in het klein” na te bootsen. Je krijgt dan de raad om dezelfde proportie mannen en vrouwen te nemen zoals ze ook in de populatie voorkomt. En je kan ook de leeftijdscategorieën imiteren en nog veel andere dingen, zoals religieuze of filosofische overtuiging, het hebben van een job, enz.. En dan moet je volgens die criteria een bepaald aantal mensen ondervragen.

Om te weten of een bepaalde buurt het beleid van de burgemeester steunt kan je de opdracht krijgen om mensen in die buurt daarover te ondervragen. Je moet daarbij 15 mannen en 20 vrouwen ondervragen. Bij de mannen moet je er 2 nemen die jonger zijn dan 20 jaar, 5 tussen de 20 en 40, enz.. Ook krijg je quota toegewezen per leeftijdscategorie van de vrouwen. De steekproef op basis van quota heeft meerdere problemen. Er is vooreerst de manier waarop je die quota bepaalt. Misschien heb je een belangrijke factor over het hoofd gezien, zoals gezinsinkomen. Een tweede probleem is dat je, binnen die quota, nog altijd de vrijheid hebt om mensen zelf te kiezen. En daar duikt het probleem van “beschikbaarheid” op. Als je naar die buurt gaat dan kom je daar bijvoorbeeld niet de mensen tegen die op dat ogenblik in de naburige stad zijn gaan werken. Bovendien spreek je gemakkelijker mensen aan die je op het eerste zicht sympathiek vindt.

3.2. Toevalsgestuurde methoden (*aselecte steekproeven*)

3.2.1. EAS Enkelvoudige aselecte steekproef (*SRS simple random sampling*)



Bij het trekken van een enkelvoudige aselecte steekproef ga je er niet alleen voor zorgen dat elk element uit de populatie dezelfde kans heeft om in de steekproef terecht te komen. Bij een EAS zorg je er ook voor dat elk mogelijk groepje van n elementen dezelfde kans heeft om jouw steekproef van grootte n te zijn.

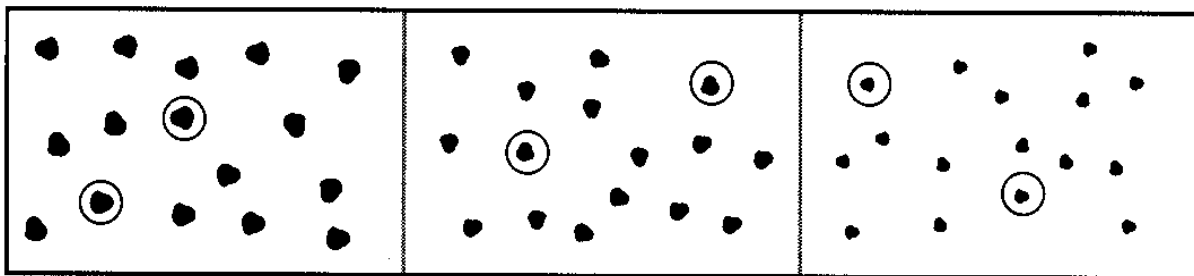
In de praktijk kan je “het toeval” best niet overlaten aan mensenhanden. Zelfs kaartjes die allemaal even groot zijn en goed door elkaar geschud kan je moeilijk volledig toevallig trekken. In feite zou je dan moeten beginnen met lukraak één kaartje te trekken (elk kaartje heeft dezelfde kans om getrokken te worden), daarna moet je de overige kaartjes weer goed schudden en dan terug een kaartje trekken (waarbij elk van de overblijvende kaartjes met een gelijke kans getrokken wordt), enz..

Het is beter om een genummerde lijst van alle elementen van je populatie te maken en dan een toevalsgenerator lukraak n verschillende getallen te laten kiezen. Als je op die manier lukraak drie leerlingen trekt uit een klas van 30, dan is het niet enkel zo dat elke leerling dezelfde kans heeft om tot jouw steekproef te behoren. Het is bovendien ook zo dat elk van de 4060 mogelijke verschillende groepjes van 3 leerlingen dezelfde kans heeft om jouw steekproef van grootte 3 te zijn. Daarom is de EAS “als methode” een goede manier om informatie over een populatie in te winnen.

Je kan voorbeelden bedenken waarbij elk element uit de populatie dezelfde kans heeft om in je steekproef te zitten maar waarbij de gebruikte methode helemaal niet goed is om de populatie te weerspiegelen. In een school met 100 jongens en 100 meisjes zou je als volgt kunnen te werk gaan. Gooi een muntstuk. Als het op kruis valt kies je lukraak (volgens een EAS) 10 jongens. Als het op munt valt kies je lukraak 10 meisjes. Dit levert je een steekproef van grootte 10 waarbij elke leerling uit die school dezelfde kans had om in je steekproef terecht te komen. Maar je uiteindelijke steekproef geeft een volledig vertekend beeld van de populatie, want er zit (als “methode” en dus niet als toeval) geen enkele jongen (of geen enkel meisje) in.

Een EAS levert “als methode” goede resultaten voor statistische verwerking. Het is de “basismethode” voor heel wat statistisch onderzoek.

3.2.2. Gestratificeerde steekproef (*Stratified random sampling*)



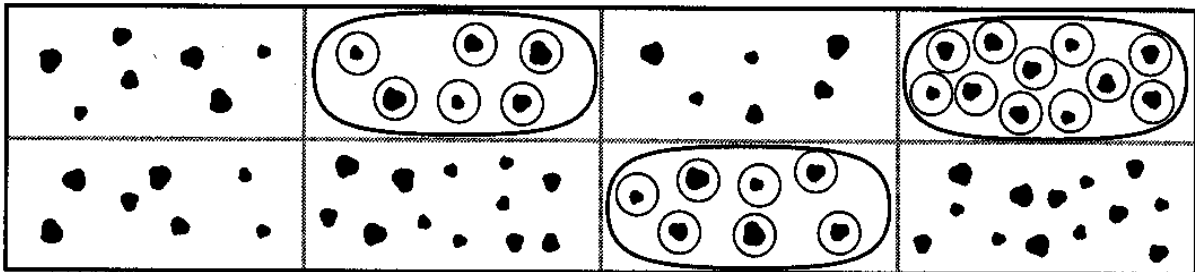
Om een gestratificeerde steekproef te trekken verdeel je eerst de populatie in niet-overlappende delen die samen de totale populatie vormen (strata). Dan trek je uit elk deel een EAS. Hierbij is het meestal goed om de steekproefgrootte proportioneel te nemen met de grootte van het stratum waaruit je trekt.

Honderd moeders hebben een reis gewonnen en mogen hun kinderen tussen 9 en 11 jaar gratis meenemen naar Disneyland. Voor de heen en terugreis wordt een chartervliegtuig ingehuurd en de organisator wil een benaderend idee hebben van het gewicht van de totale groep. Er zijn 100 moeders en 250 kinderen. Er is een groot verschil in gewicht tussen een groep volwassen vrouwen en een groep kinderen. Daarom verdeel je de populatie in 2 tamelijk homogene subgroepen (2 strata), namelijk de moeders en de kinderen. In elke subgroep trek je dan een EAS en je houdt hierbij rekening met de relatieve grootte van de groep waaruit je trekt. Als je

in totaal een steekproef van grootte 35 wil hebben, dan zal je 10 moeders en 25 kinderen nemen.

Een gestratificeerde steekproef geeft dikwijls meer precieze resultaten dan een EAS. Bij een EAS is het mogelijk dat je 35 kinderen (of 35 moeders) in je steekproef hebt. De gestratificeerde steekproefmethode beschermt je tegen zo'n situatie. Maar niets komt zo maar gratis. Om gebruik te kunnen maken van een goede stratificatie moet je vooraf reeds goede informatie hebben over karakteristieken van je populatie en over de manier waarop die karakteristieken je opmetingen kunnen beïnvloeden. Als je iets wil weten over het gewicht dan weet je vooraf dat er een groot verschil is tussen kinderen en volwassenen. Stratifiëren op basis van blond of donker haar zou hier waarschijnlijk totaal verloren moeite zijn.

3.2.3. Geclusterde steekproef (*Cluster sampling*)



Een geclusterde steekproef start met een opdeling van de volledige populatie in niet overlappende subgroepen (clusters). Hierbij is het interessant dat deze subgroepen niet homogeen zijn, maar een mengeling van elementen bevatten die de variabiliteit in de populatie weerspiegelt. Dan trek je een EAS van clusters en bij elke getrokken cluster meet je alle elementen op die in die cluster zitten.

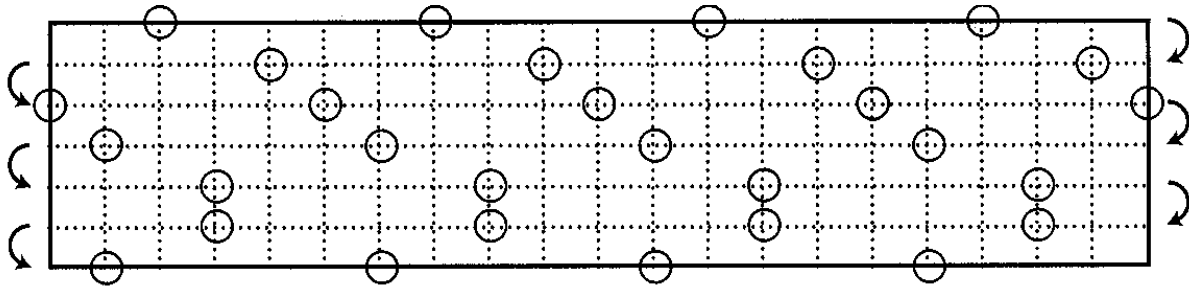
Vormt het basisonderwijs in Vlaanderen kinderen die in staat zijn om eenvoudige tabellen en grafieken uit de media te lezen en te begrijpen? Om dit na te gaan wordt een gestandaardiseerde toets opgesteld die men aan een 300-tal kinderen uit het 6^e leerjaar wil voorleggen. Voor een EAS heb je de namen van alle kinderen in het 6^e leerjaar in Vlaanderen nodig en na het trekken van de steekproef moet je heel Vlaanderen rondreizen om kind per kind die toets af te nemen (en het zijn er 300). Dit kost te veel en duurt te lang. Een alternatief is dat je de lijst neemt van alle lagere scholen in Vlaanderen (eigenlijk van alle 6^e leerjaren) en daaruit een EAS trekt van bijvoorbeeld 15 klasjes. Je hoeft nu enkel naar die 15 klasjes te rijden en daar de toets te geven aan alle kinderen in de klas.

Een geclusterde steekproef wordt dikwijls toegepast omdat zij eenvoudiger uitvoerbaar is, of omdat je geen andere informatie hebt. Je kan bijvoorbeeld wel over een lijst van clusters beschikken maar niet over een lijst van alle afzonderlijke elementen. Maar zodra je een cluster getrokken hebt dan kan het wel mogelijk zijn om daarna te achterhalen uit welke elementen hij bestaat.

In Soweto zijn heel wat kinderen niet geregistreerd en bestaan er geen precieze bevolkingslijsten. Als je daar iets over de gezondheidstoestand van baby's wil te weten komen kan je nooit aan een lijst geraken met alle namen. Je kan wel het gebied opdelen in wijken. Trek dan lukraak enkele wijken, ga naar die wijken en ondervraag huis per huis de ouders over hun baby's.

Een geclusterde steekproef levert niet altijd even precieze informatie als een EAS omdat het mogelijk is dat elementen binnen eenzelfde cluster meer op elkaar lijken. Je krijgt dan te veel keer een analoge opmeting. Veel hangt ook af van welke eigenschap je onderzoekt.

3.2.4. Systematische steekproef (*Systematic sampling*)



Bij een systematische steekproef kies je systematisch elke “zoveelste” persoon (bijvoorbeeld elke tiende persoon) op een lijst. Om het toeval hierbij een rol te laten spelen kies je het startpunt lukraak. Je laat bijvoorbeeld een toevalsgenerator een getal tussen 1 en 10 kiezen. Als dat een 3 is dan start je op de lijst met de derde persoon. Daarna neem je de 13^e, de 23^e, enz..

Als er geen enkel verband is tussen de volgorde op de lijst en de grootte die je opmeet (maar hoe kan je dat weten?) dan kan een systematische steekproef resultaten opleveren die bepaalde eigenschappen van de populatie behoorlijk weerspiegelen. Maar echte houvast heb je hier niet.

Een systematische steekproef is eigenlijk een speciaal geval van een geclusterde steekproef. Je hebt de populatie opgedeeld in clusters, en voor deze opdeling heb je je moeten houden aan de volgorde van de vooraf opgemaakte lijst. Zo heb je (als je stappen van 10 gekozen hebt) een eerste cluster met nummers {1, 11, 21, 31,...}. De andere clusters zijn {2, 12, 22, 32,...}, {3, 13, 23, 33,...}, enz. Door nu lukraak het startpunt te kiezen kies je eigenlijk lukraak één cluster waarna je alle elementen uit die cluster opmeet.

De EAS is de basis voor de toevalsgestuurde steekproefmethoden.

Andere toevalsgestuurde methoden zijn soms handiger of preciezer maar dikwijls ook ingewikkelder.

In ieder geval moet elke toevalsgestuurde selectiemethode voldoen aan de volgende eigenschappen:

1. de onderzoekers hebben helemaal niets te zeggen over welke elementen er precies in de steekproef terechtkomen (geen keuze gebaseerd op persoonlijk oordeel, op eigen ervaring, op zogenaamd gezond verstand, of op wat dan ook)
2. er is een vooraf vastgelegde procedure om de steekproef te selecteren, met een gekende manier over hoe het toeval daarbij zijn rol speelt.

Het gevolg hiervan is dat het bij toevalsgestuurde methoden mogelijk is om van elk individu op de steekproeflijst de kans te berekenen dat het in de steekproef terecht komt. Dit is een minimale vereiste voor elk statistisch onderzoek.

3.2.5. Getrapte steekproef (*multi-stage sampling*)

Een getrapte steekproef is niets nieuws. “Getrapd” is enkel een woord waarmee je aangeeft dat de methode om aan je steekproef te geraken bestaat uit een combinatie van steekproefmethoden. Je kan meerdere keren dezelfde methode na elkaar toepassen of je kan verschillende methoden combineren.

Als je leerlingen uit Vlaamse scholen wil opmeten, dan kan je beginnen met een lijst van scholen in plaats van met een lijst van leerlingen. Uit de lijst van scholen trek je dan bijvoorbeeld lukraak 20 scholen. In plaats van alle leerlingen in die 20 scholen op te meten kan je daarna per school een lijst maken van alle klassen en per school lukraak 3 klassen kiezen. En bij die gekozen klassen ga je alle leerlingen opmeten. Dit betekent dat je een eerste geclusterde steekproef neemt waarbij de clusters gevormd worden door de scholen. Bij elke geselecteerde cluster (school) maak je nieuwe clusters door alle klassen van die school te bekijken en dan lukraak een aantal klassen te kiezen. En dan meet je alle elementen (leerlingen) op van de geselecteerde klassen. In dit voorbeeld gebruik je twee keer na elkaar de methode van de geclusterde steekproef.

Je kan natuurlijk ook 20 scholen lukraak kiezen en dan per school stratifiëren op jongens en meisjes. Per school trek je dan een EAS uit de jongens en een EAS uit de meisjes, met steekproefgroottes die proportioneel zijn met het aantal jongens en meisjes in die school. De steekproef die je uiteindelijk bekomt is ontstaan door eerst een geclusterde steekproef te trekken en dan bij de geselecteerde clusters (= geselecteerde scholen) een gestratificeerde steekproef te trekken.

Grote statistische onderzoeken gebruiken meestal een of andere combinatie van steekproefmethoden en werken dus met getrapte steekproeven. De theorie om op basis van getrapte steekproeven wetenschappelijk verantwoorde uitspraken te doen in de verklarende statistiek is veel ingewikkelder dan wanneer je met een EAS werkt.

4. Methoden om eigenschappen van steekproefelementen op te meten

4.1. De “rechtstreekse” observatie of meting

Om zo weinig mogelijk fouten te maken zou het ideaal zijn dat de onderzoeker zelf alle elementen uit de steekproef kan opmeten en dat hij dit doet op een gestandaardiseerde manier en met geijkte meetinstrumenten. Denk bijvoorbeeld aan een onderzoeker die bij de individuen in zijn steekproef de leeftijd en het geslacht noteert en met één en dezelfde geijkte weegschaal hun gewicht bepaalt. In heel wat grote studies is het niet zo eenvoudig en moeten meerdere medewerkers getraind worden om op eenzelfde manier karakteristieken van de steekproefelementen op te meten. Bij rechtstreekse observatie of meting hoeft het natuurlijk niet over individuen te gaan. Je kan allerlei soorten “elementen” in je steekproef hebben waarbij je de waarde van bepaalde “veranderlijken” opmeet, zoals het gewicht van eieren of de hoogte van bomen.

4.2. De enquête (survey)

Een enquête is een typisch meetinstrument om bij mensen allerlei informatie in te winnen. Die informatie kan te maken hebben met eigenschappen die “objectief” waarneembaar zijn (zoals leeftijd, lengte en gewicht). Het kan ook over eerder “subjectieve” kenmerken gaan (zoals de mate waarin je je misselijk voelt na narcose bij een operatie). Bij een “enquête die gebaseerd is op een toevalsgestuurde steekproef” (*sample survey*) meet je eigenschappen op van elementen die op die toevalsgestuurde manier in je steekproef zijn terechtgekomen. Een census (*census*) is een enquête waarbij je de hele populatie opmeet.

Wat kan er allemaal fout gaan? Soorten vertekening.

Elke steekproefmethode waarbij je resultaten bekomt die op een systematische manier falen om de populatie te weerspiegelen is, als methode, “vertekend” (*biased*).

Als een steekproefmethode vertekend is dan helpt het niet om een grotere steekproef te trekken want dan herhaal je de fundamentele fout op een nog grotere schaal.

Of data volgens een goede steekproefmethode verzameld zijn kan je meestal niet te weten komen door alleen maar naar die data te kijken. Je moet expliciet vragen aan de onderzoeker op welke manier de steekproef getrokken is. Als je zelf een statistisch rapport schrijft dan is het belangrijk dat je duidelijk uitlegt hoe je de data verzameld hebt.

5. Vertekening door de manier waarop je aan de elementen van je steekproef geraakt (*sample selection bias*)

5.1. Vertekening door de selectiemethode die niet “neutraal” is (niet gebaseerd op een vooraf gekend toevalsmechanisme)

5.1.1. Steekproef op basis van vrijwillige reactie

In een grote krant vroeg columniste Ann Landers aan ouders of zij terug kinderen zouden willen als ze de kans kregen om het over te doen. Landers kreeg meer dan 10 000 antwoorden. Een overweldigende meerderheid (70%) antwoordde negatief en schreef dat kinderen de moeite niet waard zijn. Een goed opgezette statistische studie toonde achteraf aan dat 90% van de ouders tevreden zijn met hun beslissing om kinderen te hebben. Wat zou een oorzaak kunnen zijn van het grote verschil in deze twee studies? Welke ouders schrijven spontaan naar een krant om op die vraag te antwoorden?

5.1.2. Steekproef op basis van bereikbaarheid

In 1936 liep de eerste termijn van het presidentschap van F.D. Roosevelt ten einde en waren er opnieuw verkiezingen. De tegenkandidaat van (democraat) Roosevelt was de republikeinse gouverneur A. Landon uit Texas. De meeste waarnemers dachten dat Roosevelt voor een tweede ambtstermijn zou herkozen worden, maar de enquête van het magazine "*Literary Digest*" voorspelde iets helemaal anders. Op basis van de 2.4 miljoen antwoorden die waren binnengekomen (het grootste aantal mensen in de geschiedenis dat ooit op een enquête heeft geantwoord) voorspelden zij dat Landon zou winnen met 57% van de stemmen tegenover slechts 43% voor Roosevelt. Maar de verkiezingen draaiden uit op een klinkende overwinning voor Roosevelt, met 62% tegenover 38%.

Hoe kon de "*Literary Digest*" (die kort nadien failliet ging) op basis van zo'n reuzengrote enquête toch nog die enorme fout maken? Het antwoord zit zowel in de steekproef op basis van bereikbaarheid als in de hoge non-respons. De "*Literary Digest*" had namelijk 10 miljoen formulieren per post verstuurd en had daarvoor adressen gebruikt van hun lezerslijst maar ook uit telefoonboeken enz.. In die tijd had 3/4 van de mensen geen telefoon, en was de armere groep van de bevolking zeker niet geabonneerd op de "*Literary Digest*". Er werden dus bepaalde groepen van mogelijke kiezers systematisch uitgesloten. Verder namen slechts 20% van de aangeschrevenen de moeite om dat formulier in te vullen en terug te sturen. En het waren natuurlijk de ontevredenen (die boos waren op Roosevelt en een andere president wilden, namelijk Landon) die overweldigend antwoordden. De mensen die tevreden waren met de gang van zaken en vonden dat Roosevelt nog rustig een tweede termijn kon krijgen (en eigenlijk verwachtten dat dit ook wel zou lukken) namen meestal de moeite niet om die enquête terug te sturen.

De conclusie van deze enquête was dus totaal waardeloos.

George Gallup, die toen juist een eigen enquêtebureau oprichtte, gebruikte toevalsgestuurde methoden en voorspelde op basis van de ondervraging van 50 000 Amerikanen dat Roosevelt zou winnen. Hij deed zelfs straffer. Hij voorspelde de (foutieve) voorspelling van de *Literary Digest*, nog vooraleer het magazine zijn eigen resultaten binnen had en zijn conclusie kon publiceren. Daarvoor had hij gewoon een lukrake steekproef van grootte 3 000 genomen uit de lijst van 10 miljoen adressen die de *Literary Digest* had gebruikt.

5.1.3. Steekproef op basis van quota

In 1948 was T. Dewey de republikeinse tegenkandidaat bij de presidentsverkiezingen waarbij de democratische president H. Truman hoopte herverkozen te worden. Alle polls (inclusief Gallup) beweerden dat Dewey zou winnen. Het draaide echter anders uit en Harry Truman werd herkozen als president. Wat was er fout gelopen?

Elk enquêtebureau had het trekken van zijn steekproef uitgevoerd op basis van quota. Een medewerker bij Gallup bijvoorbeeld moest in de stad St. Louis 13 personen ondervragen en kreeg volgende instructie. Neem 7 mensen die in de binnenstad wonen en 6 uit de buitenwijken. Zorg ervoor dat je 6 vrouwen en 7 mannen hebt. Bij de 7 mannen moeten er 3 jonger zijn dan 40 jaar en 4 ouder dan 40. Je moet 6 blanken en 1 zwarte hebben. Zorg ervoor dat er 2 zijn met een laag inkomen, 3 uit de middenklasse en 1 rijke (de bedragen waren

aangegeven). Op eenzelfde manier werd ook gezegd hoe de 6 vrouwen moesten gekozen worden.

Deze karakteristieken waren uitgezocht omdat zij belangrijk waren voor het stemgedrag, en dat waren zij ook. Maar er zijn er nog veel meer, en je kan gewoon niet weten waarvan het allemaal afhangt dat iemand Republikein of Democraat stemt.

Een nog groter probleem bij quotasteekproeven is het feit dat de onderzoeker volgens eigen oordeel mensen mag aanspreken, zolang hij maar binnen zijn quota blijft. In 1948 waren mensen die Republikein stemden gemiddeld iets welstellender. Zij waren hoger geschoold, hadden meestal telefoon, woonden in mooiere en veiliger buurten. Kortom, zij waren gemakkelijker om te contacteren. Wat blijkbaar ook (veel te veel) is gebeurd, binnen de opgelegde quota!

5.2. Vertekening door een onvolledige of een verkeerde steekproeflijst

5.2.1. Ondervertegenwoordiging (*undercoverage*)

Zelfs als je niet zomaar op basis van beschikbare lijsten werkt (gemakkelijke bereikbaarheid) is het in de praktijk dikwijls moeilijk om methoden te bedenken die werkbaar zijn en waarbij iedereen van de bestudeerde populatie in de steekproef kan terechtkomen. Om het financieel haalbaar te houden gebeuren heel wat enquêtes telefonisch. Als je daarvoor een gewone telefoongids gebruikt dan mis je zeker al de groep van mensen die geen telefoon heeft. Maar zelfs als je studie eruit bestaat dat je aan mensen met telefoon vraagt of zij tevreden zijn over de diensten van hun operator heb je een probleem. Er zijn immers nogal wat mensen met een “geheim” nummer, en die zijn als groep misschien wel verschillend van de “doorsnee” telefoonabonnee. Enquêtebureaus laten soms door een computer lukraak nummers genereren, en bellen die dan op (*random digit dialing*). Zo krijgen zij ook geheime nummers te pakken. Maar zelfs dan zijn er groepen ondervertegenwoordigd. Topfunctionarissen nemen nooit zelf de telefoon op maar laten dat hun secretaresse doen, en die schakelt nooit een gesprek door dat begint met “hallo, dit is een enquête, mogen we enkele vragen stellen?”.

5.2.2. Vertekening door omvang (*size bias*)

Finland wordt niet voor niets “het land der duizend meren” genoemd. Een bioloog wil een idee hebben over de vissoorten in de Finse meren en zal daarvoor 45 meren ter plaatse gaan onderzoeken. Hij beschikt over een zeer grote landkaart van Finland waarop alle meren zijn aangeduid. Hij legt die kaart op de grond, neemt een rijstkorrel en gooit die lukraak op de kaart. Als de rijstkorrel op een meer valt dan noteert hij daarvan de ligging. Hij herhaalt deze werkmethode tot hij op die manier 45 verschillende meren geselecteerd heeft. Een collega van deze bioloog vindt rijst gooien niet wetenschappelijk verantwoord. Hij beschikt over een computerprogramma dat volledig lukraak punten op een kaart kan kiezen. Op basis van die lukrake punten kiest deze collega zijn 45 meren.

Je kan denken dat die collega een EAS uit de populatie van alle Finse meren getrokken heeft, maar dat is niet waar. Hij heeft lukraak punten op een landkaart aangeduid en zo’n punt heeft 10 keer meer kans om terecht te komen op een meer dat 10 keer zo groot is als een ander meer. Er is niet gewerkt met een lijst met namen van de meren maar met een grafische voorstelling waarbij grotere meren een grotere oppervlakte op de landkaart innemen. Lukraak

een punt aanduiden op zo'n landkaart resulteert in een steekproef met vertekening door omvang, waarbij "omvang" in dit voorbeeld verwijst naar "oppervlakte".

In de wintermaanden zijn er in het eerste leerjaar nogal wat leerlingen afwezig wegens ziekte. Een directrice beweert dat 20 jaar geleden de zieke kinderen veel langer thuis konden blijven om rustig te herstellen. "Tegenwoordig worden ze veel sneller terug naar school gestuurd" beweert zij, "en dat heeft te maken met ouders die allebei uit werken gaan". Zij wil haar bewering staven met cijfers op basis van een steekproef. Het zou een reuzenwerk zijn om in de archieven al de kinderen te gaan opzoeken die destijds ziek geweest zijn. Daarom kiest zij een lukrake steekproef van 6 dagen uit de schooldagen van de maanden december, januari en februari van 20 jaar geleden. In het archief wordt dan alleen voor die 6 dagen opgezocht welke kinderen afwezig waren en hoelang die afwezigheid duurde. Een berekening van het gemiddelde geeft de directrice gelijk. De zieke kinderen werden destijds gemiddeld langer thuisgehouden.

Je kan de indruk hebben dat een EAS trekken uit de lijst van de schooldagen en dan de afwezige kinderen noteren op hetzelfde neerkomt als een EAS trekken uit alle kinderen die toen ziek geweest zijn (voor de eenvoud mag je zelfs onderstellen dat elk kind niet meer dan één keer ziek was in die winter). Op onderstaand vereenvoudigd voorbeeld, dat over één week gaat, kan je zien dat dit niet zo is. Van alle kinderen in die klas zijn er 5 die elk één dag ziek zijn en er is één kind dat de hele week ziek is. Per ziek kind is er dus een gemiddelde afwezigheid van $10/6 = 1.7$ dagen.

	maandag	dinsdag	woensdag	donderdag	vrijdag
ziek kind 1					
ziek kind 2					
ziek kind 3					
ziek kind 4					
ziek kind 5					
ziek kind 6					

Trek nu lukraak één dag. In je steekproef heb je dan één kind dat 5 dagen ziek is en één kind dat 1 dag ziek is. Gemiddeld is dat $6/2 = 3$ dagen afwezigheid per ziek kind. Je steekproefmethode heeft als resultaat dat kinderen die langer ziek zijn meer kans hebben om in je steekproef terecht te komen. Je hebt hier vertekening door omvang, waarbij "omvang" in dit voorbeeld verwijst naar "tijdsduur".

5.3. Vertekening door ontbrekende respons

5.3.1. Ontbrekende informatie bij directe opmetingen (*missing data*)

Veel onderzoeken hebben ontbrekende gegevens. Dat kan zelfs over de meest eenvoudige dingen gaan, die heel gemakkelijk te noteren zijn. Voor een studie van te vroeg geboren kinderen werden 1200 medische dossiers opgemaakt met allerlei gegevens over gewicht, bloedgroep, enz.. Bij die dossiers waren er zeven waarbij de dokter (of de verpleegster) vergeten was om het geslacht in te vullen!

Als er bij de opmetingen van een veranderlijke weinig ontbrekende gegevens zijn en als je mag onderstellen dat die ontbrekende gegevens niets te maken hebben met de studie zelf, dan

heb je geen groot probleem. Maar bij een groot aantal ontbrekende gegevens moet je op je hoede zijn en op zoek gaan naar mogelijke verklaringen. Als je alleen maar werkt met de gegevens die werden opgemeten en de ontbrekende gegevens gewoon negeert kan je wel eens eindigen met een totaal verkeerde conclusie.

Bij het uittesten van een hogere dosis voor de behandeling van een bepaalde aandoening werd aan de patiënten gevraagd om die hogere dosis gedurende een week elke dag in te nemen. Een verpleegster zou dan het effect noteren dat teweeg gebracht wordt door een week lang zo'n hogere dosis in te nemen. Voor sommige patiënten was er duidelijk verbetering en zij waren blij dat zij elke dag het nieuwe geneesmiddel konden nemen. Bij anderen veroorzaakte die hogere dosis maagkrampen en duizeligheid. Die patiënten hielden het hoogstens een dag of twee uit en beslisten dan om te stoppen met die hogere dosis. Het eindresultaat was dat er heel wat dossiers waren waarbij er geen antwoord was op de vraag: "hoe reageert een patiënt die een week lang elke dag deze nieuwe medicatie neemt?". Bij de volledige dossiers echter was te lezen dat de hogere dosis prima werkt.

5.3.2. Non-respons bij enquêtes (*non-response bias*)

Als bij enquêtes een grote groep mensen het formulier niet terugstuurt dan heb je een probleem. De kans is groot dat de groep der non-respondenten serieus verschilt van de groep der respondenten. Bij een enquête per brief proberen enquêtebureaus de respons op te drijven door tot twee keer toe een herinnering te sturen. Andere systemen bestaan erin om enquêteurs te trainen om mensen te gaan ondervragen in een persoonlijk interview. Ook hier worden bepaalde categorieën gemist. Sommige mensen zijn systematisch niet aanwezig op de uren waarop enquêteurs langskomen. Analoge problemen heb je bij de telefonische enquête. Bovendien zijn er, bij elke soort enquête, mensen die niet antwoorden. Misschien vormen die juist een speciale subgroep van de te onderzoeken populatie.

6. Vertekening door de manier waarop je informatie opmeet of inwint (*response bias*)

6.1. Vertekening door het vragenformulier van de enquête (*questionnaire bias*)

6.1.1. Vragen die onnauwkeurig of te ingewikkeld zijn (*wording of questions*)

Op 20 April 1993 blokletterde de Los Angeles Times dat een enquête uitwees dat 1 op 3 Amerikanen eraan twijfelen dat de Holocaust ooit heeft plaatsgevonden. De vraag in die enquête was als volgt: "Does it seem possible or does it seem impossible to you that the Nazi extermination of the Jews never happened?". Een jaar later werd dezelfde enquête terug afgenomen. Alle vragen bleven onveranderd behalve de vraag over de Holocaust. Die was nu als volgt geformuleerd. "Does it seem possible to you that the Nazi extermination of the Jews never happened, or do you feel certain that it happened?". Op deze vraag antwoordde maar 1% dat het mogelijk was dat dit nooit was gebeurd en 8% zei dat ze het niet wisten.

Zelfs de volgorde kan een invloed hebben.

Je krijgt andere antwoorden op de vraag “Is het verkeer een grotere luchtvervuiler dan de industrie?” dan op de vraag “Is de industrie een grotere luchtvervuiler dan het verkeer?”

Tijdens de periode van de koude oorlog tussen Rusland en Amerika liep er een onderzoek in Amerika waarbij de volgende twee vragen in een enquête voorkwamen:

- A. Denk jij dat Amerika moet toelaten dat Russische reporters naar Amerika komen en dat zij hun nieuwsverslagen ongecensureerd naar Rusland mogen sturen?
- B. Denk jij dat Rusland moet toelaten dat Amerikaanse reporters naar Rusland gaan en dat zij hun nieuwsverslagen ongecensureerd naar Amerika mogen sturen?

Het antwoord op vraag A. was “JA” bij 36 % van de respondenten wanneer die vraag als eerste stond. Niet minder dan 73 % van de respondenten antwoordden “JA” wanneer vraag A. als tweede stond.

6.1.2. Vragen die gevoelig liggen (*sensitive questions*)

Mensen geven zij niet graag toe dat zij onwettelijk handelen (belastingen ontduiken, in het zwart werken, drugs dealen ...). Omgekeerd wil men zich soms beter voordoen dan men is. Bij een enquête die peilde naar de betrokkenheid tegenover mensen die het minder goed hebben werd gevraagd hoeveel men jaarlijks besteedde aan “het goede doel”. Er werden onwaarschijnlijk hoge bedragen vermeld. Voor vragen die als zeer “gevoelig” of “bedreigend” overkomen bestaan speciale enquêtemethodes (waarbij anonimiteit op een extra manier gewaarborgd wordt).

6.2. **Vertekening door foutief antwoord of door slecht geijkt meetinstrument**

6.2.1. Systematische meetfouten (*measurement bias*)

Een onnauwkeurig geijkt meetinstrument (weegschaal, chronometer, ...) geeft bovenop de variabiliteit van de opmetingen ook nog eens een systematische vertekening (bijvoorbeeld systematisch een gram te weinig).

6.2.2. Verkeerd antwoord (*incorrect response bias*)

De meeste mensen die toestemmen om persoonlijk geïnterviewd te worden willen niet onvriendelijk overkomen en proberen soms te gissen welk antwoord de enquêteur graag zou horen. Dikwijls spelen zij daar (onbewust) op in. Veel verkeerde antwoorden komen gewoon ook voort uit het feit dat mensen zich dingen niet meer zo goed herinneren, zelfs al denken zij dat ze het toch nog juist weten.